DOCUMENT RESUME

ED 438 925 PS 028 343

AUTHOR	St. Pierre, Robert G.; Ricciuti, Anne; Creps, Cynthia
TITLE	Synthesis of Local and State Even Start Evaluations, 2000.
INSTITUTION	Abt Associates, Inc., Bethesda, MD.
SPONS AGENCY	Department of Education, Washington, DC. Office of the Under Secretary.
REPORT NO	OUS-PES-2000-02
PUB DATE	2000-00-00
NOTE	49p.; For the 1994-1997 final report of the national evaluation of Even Start, see ED 427 889. For the 1998 report of Even Start, see ED 427 890.
CONTRACT	43-31KV-7-A2049
AVAILABLE FROM	ED Pubs, P.O. Box 1398, Jessup, MD 20794-1398. Tel: 800-872-5327 (Toll Free); Fax: 301-470-1244; e-mail: edpubs@inet.ed.gov; Web site: http://www.ed.gov/pubs/edpubs.html; For full text: http://www.ed.gov/pubs/evenstart final/synthesis/. edpubs@inet.ed.gov (Document No. 2000-02).
PUB TYPE	Reports - Evaluative (142)
EDRS PRICE	MF01/PC02 Plus Postage.
DESCRIPTORS	Demonstration Programs; Early Childhood Education; *Early Intervention; Evaluation Research; *Family Literacy; Models; Program Descriptions; Program Effectiveness; *Program Evaluation; Program Improvement; Young Children
IDENTIFIERS	*Even Start; Outcome Oriented Evaluation

ABSTRACT

The Even Start program, designed to demonstrate ways of implementing and delivering family literacy services, has had strong evaluation requirements since its inception. An analysis was conducted to determine the types of evaluations programs had conducted and their findings concerning Even Start outcomes. A convenience sample of 4 state evaluations and 118 local evaluations from 19 states were used for the analysis. Findings indicated that programs rarely used data for program management and improvement. Almost all evaluations reported some information on project implementation, including descriptions of project structure and activities, recruiting and retaining families, and working with collaborative agencies. More than three-quarters of the evaluations contained information about program outcomes, although none incorporated the use of control or equivalent comparison families. Twenty-four evaluations were selected for detailed examination of outcomes. These evaluations showed consistent positive gains on measures of child and adult development. Half the reports gauged the size of outcomes by making comparisons to national data, to evaluations of other early childhood or adult education programs, or to normal development. Because local studies did not provide much useful information to local project directors, State Coordinators, or to the Department of Education, several recommendations were made, including changes in Even Start legislation language, provision of guidance regarding the amount of funding needed for conducting evaluations focusing on program improvement, and the establishment of a community of local evaluators. (Includes a list of principles for the continuous improvement approach.) (KB)

Reproductions supplied by EDRS are the best that can be made
from the original document.

ED 438 925

PES

PES

PLANNING AND EVALUATION SERVICE

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
X This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

SYNTHESIS OF LOCAL AND STATE EVEN START EVALUATIONS

2000

PS 028343

BEST COPY AVAILABLE

U.S. DEPARTMENT OF EDUCATION OFFICE OF THE UNDER SECRETARY

DOC #2000-02

Synthesis of Local and State Even Start Evaluations

2000

Prepared for
U.S. Department of Education
Office of the Under Secretary
Planning & Evaluation Service

Prepared by
Robert G. St.Pierre
Anne Ricciuti
Cynthia Creps
Abt Associates Inc.

Permission to reprint this publication is not necessary, but the citation should be: U.S. Department of Education, Office of the Under Secretary, *Synthesis of Local and State Even Start Evaluations*, Washington, D.C., 2000.

This report is also available on the Department's web site at: www.ed.gov/offices/OUS/eval/elem.html.

This report was prepared under purchase order to Abt Associates Inc., purchase order number 43-31KV-7-A2049. The views expressed in this report do not necessarily reflect the position or policy of the Department, and no official endorsement by the U.S. Department of Education is intended or should be inferred.

On request, this publication is available in alternate formats, such as braille, large print, audiotape, or computer diskette. For more information, please contact the Department's Alternative Format Center at (202) 205-8113.

Copies of this report are available by contacting the U.S. Department of Education's Publication Center in the following ways: Toll-free phone calls to 1-877-4ED-Pubs (1-877-433-7827), TTY/TDD call 1-877-576-7734. If 877 is not yet available in your area, call 1-800-USA-LEARN (1-800-872-5327), TTY/TDD call 1-800-437-0833; via internet at http://www.ed.gov/pubs/edpubs.html; via e-mail at edpubs@inet.ed.gov; via fax to 301-470-1244; and, via mail to ED Pubs, Education Publications Center, U.S. Department of Education, P.O. Box 1398, Jessup, MD 20794-1398.

Table of Contents

Executive Summary 1
- Even Start Evaluation Requirements 1
- Obtaining State and Local Evaluation Reports 2
- Using Data for Program Improvement 2
- Local Evaluation Methods 3
- Evaluation Findings 3
- Recommendations 5

1.0 Overview of Even Start 9
- Program Operations 9
- Design of Even Start Projects 11
- Variations on the Basic Model 13

2.0 Even Start Evaluation Requirements 15
- National Evaluation 15
- State Evaluation 15
- Local Evaluation 16

3.0 Study Methods 21
- Research Questions 21
- Obtaining Evaluation Reports 21
- Receiving and Logging-In Reports 22
- Describing Methods Used to Conduct Local Evaluations 22
- Describing Findings from a Subset of Evaluations 23

4.0 Description of Local and State Evaluations 25
- Using Local and State Evaluations for Program Improvement 25
- Quality of Local and State Evaluations 27
- Description of Implementation Studies 28
- Description of Outcome Studies 29

5.0 Evaluation Review 35
- Outcomes for Children 35
- Outcomes for Adults 37
- Impacts on Children and Adults 39
- Participant Characteristics Associated with Positive Outcomes 41

6.0 Recommendations 43

7.0 References 47

Executive Summary[1]

Since its inception, the Even Start program has undergone evaluative scrutiny at a variety of levels. In addition to the ongoing federally-funded national evaluation, each local project is required by law to conduct an independent local evaluation, and in some states, projects participate in a state-wide evaluation. While the primary purpose of many state and local evaluations is to help improve and focus the program in specific locales, these evaluations may contain information that could be more widely disseminated to program developers and policy makers.

To understand what types of evaluation activities are being undertaken, and to see whether state and local evaluations do, indeed, contain information that would be useful for project improvement and helpful outside of their immediate applications, the U.S. Department of Education (hereafter, the Department) asked Abt Associates Inc. to collect, read, and report on state and local Even Start evaluations. This document is the final report from that project.

This project aimed to gather and review state and local Even Start evaluations, to describe the types of evaluations that were conducted, to describe findings about the outcomes of Even Start programs, and to provide guidance on improving state and local evaluations. In practice, the study focused on local evaluations since few state-level studies were available for review at the time that we were collecting evaluation documents.

Even Start Evaluation Requirements

Even Start was initiated as a demonstration program. After its first decade, it remains a program that is designed to help demonstrate and learn about the best ways of implementing and delivering family literacy services. To assist with this mission, strong evaluation requirements have been a prominent and continuing

---

1 This report was prepared by Robert St.Pierre, Anne Ricciuti, and Cynthia Creps at Abt Associates Inc. under purchase order 43-31KV-7-A2049 with the U.S. Department of Education. The views expressed in this report do not necessarily reflect the position or policy of the Department, and no official endorsement by the Department should be inferred. Staff at the Planning and Evaluation Service within the Department were responsible for providing technical input and oversight of the study. As Project Officer for this study, Tracy Rimdzius oversaw all planning, implementation, and reporting activities. Valuable comments on early drafts of this report were provided by Tracy Rimdzius, Valena Plisko, and Patricia McKee.

feature of the Even Start legislation. In addition to the ongoing national evaluation, the legislation calls for each project to "...provide for an independent evaluation of the program." (P.L. 103-382, Sec. 1205). The only additional guidance on local evaluations comes from the Department:

> "Each Even Start project is required to provide for an independent evaluation of the project. See section 1205 (10) of the Act. These evaluations provide local projects, States, the Department, and the Congress with objective data about the activities and services provided by the project, the participants served, the retention rates of those participants, and the success of the families in the project." (Department of Education preliminary guidance).

Obtaining State and Local Evaluation Reports

During the 1996-97 school year there were about 650 local Even Start projects, and the law specifies that each project is to conduct a local evaluation. For the present study, we hoped to collect at least 100 local and state evaluation reports, though we had no idea how many reports could actually be obtained. Because some State Coordinators require their grantees to submit copies of local evaluations, we sent a letter to all State Coordinators, describing the planned synthesis and asking for their help in obtaining copies of state and local studies. Obtaining local evaluations was difficult. Having limited time and funds, we ended up with a sample of convenience – a total of 4 state evaluations and 118 local evaluations from 19 states.

Using Data for Program Improvement

Local evaluations are conducted for multiple purposes. First and foremost, Even Start projects have to comply with the legislative requirement for a local evaluation. Given this requirement, many local evaluations choose to focus on project outcomes in an attempt to provide evidence so that local project directors can obtain political and/or financial support for the program from school boards, civic organizations, corporate sponsors, or foundations.

The evidence contained in the reports that we reviewed shows that Even Start projects rarely engage in the systematic use of data to manage and improve their programs. Instead, program improvements/alterations typically are made on the basis of anecdotal evidence obtained through observations and stories gathered from the personal experiences of program implementers. A few reports that we

reviewed noted recommendations from the previous year, described whether they had been addressed, and provided additional recommendations for the current year.

Local Evaluation Methods

Of the 122 evaluations reviewed for this study, almost all reported some information on the implementation of the projects being studied including descriptions of project structure and activities, level of participation, recruiting and retaining families, and working with collaborating agencies.

More than three-quarters of the evaluations we reviewed contained information about program outcomes. The most common design for local evaluations, used in 76% of the cases,[2] was to pretest and posttest Even Start families at the start and end of a school year. No control or comparison families were included in these studies. In 31% of the outcome studies Even Start families were not pretested, but they were posttested at the end of a school year. No control or comparison families were included in these studies. Only 10% of the local outcome evaluations used a quasi-experimental design in which the gains of Even Start families were compared to the gains of families in a non-equivalent comparison group, e.g., families in a parallel program, children in Head Start, etc. Finally, no local outcome evaluation used an experimental design in which families eligible for Even Start are randomly assigned to participate in the program or in a control group.

Evaluation Findings

We selected 24 of the 122 evaluations that focused on outcomes and read them carefully to understand their findings.

Outcomes for children. Local evaluations showed a consistent pattern of positive gains on several different measures of child development including the PreSchool Inventory (PSI), PreSchool Language Scale-3 (PLS), Peabody Picture Vocabulary Test (PPVT), pre-IDEA Proficiency Test, and the High/Scope Classroom Observation Record (COR). When significance testing was done, the gains generally were statistically significant. The size of the gains is consistent with the size of the gains observed in the national Even Start evaluation and in other family literacy studies which administered these same measures.

---

2 Percentages add to more than 100 because some evaluations used more than one design.

Outcomes for adults. As was the case for child-level outcomes, local evaluations showed a consistent pattern of positive gains on several different measures of adult development including the Comprehensive Adult Student Assessment System (CASAS), the IDEA Proficiency Test, and the Tests of Adult Basic Education (TABE). Statistically significant gains were observed in almost all projects that administered these measures. In addition, most projects that collected data on GED status reported that substantial numbers or percentages of adults attained a GED during their participation in Even Start.

Estimating program impacts. Half of the reports that we reviewed gauged the size of these outcomes by making comparisons to other studies. For example:

- ***Comparison to national data.*** Six studies compared local gains on the PSI, PLS, CASAS, and the Home Screening Questionnaire (HSQ) to average gains reported in the national Even Start evaluation. As would be expected, the gains in the 6 studies were larger than national gains in some cases and smaller in others.

- ***Comparison to national statistics.*** One project found that compared to data presented in the National Education Goals Reports, Even Start parents are more likely to read to their children than parents nationally.

- ***Comparison to gains observed in other evaluations.*** Some studies compared local Even Start gains to gains from other studies of adult education or early childhood education programs. For example, one project compared Even Start parents with parents in a local JOBS program on expectations of their child and view of parenting. No differences were found between the two groups of parents. Two projects found that the CASAS gains of their adults were greater than the gains of adults in California adult education programs. Yet another program compared the CASAS gains of their adults to those of adults in a Toyota-funded family literacy project. Finally, one program found that the PSI scores of Even Start children were lower than the scores of children in a Home Start evaluation conducted 20 years ago.

- ***Comparison to "normal" development.*** Some studies compared local Even Start gains to normal development based on the use of local or national norms. For example, one project found large gains when comparing developmental age with chronological age using PPVT scores; a second project projected normal growth based on pretest scores of the

Learning Accomplishment Profile and found that Even Start children gained more than expected.

These local evaluations made an important attempt to judge the magnitude of the gains observed for Even Start children or adults. However, none of the studies was able to estimate the "true" impact of Even Start, which would require randomly assigning incoming families to Even Start or to a control group. Given the resources available to local evaluators, it is unreasonable to expect that any local study could actually put a randomized experiment in place.

Recommendations

The Department currently spends much more on local Even Start evaluations (more than $3 million per year) than on the national evaluation ($500,000 to $750,000 per year). If the mere existence of a local evaluation report is useful, or if it is useful for project directors to have a local evaluation that concludes that Even Start is a success, then local evaluations can be judged helpful to local projects. However, if the criterion for success is that a local evaluation should be used to help improve an Even Start project, then this review of more than 100 evaluations was not able to show that local studies provide much useful information, either to local project directors, to State Coordinators, or to the Department.

How can we improve this situation? Most important, in our view, is to provide clear guidance on the purpose of local evaluations. The current mix of guidance (to assess program outcomes, to address questions of local importance, to improve program functioning) needs to be streamlined and focused. We agree with the importance of the Department's recent move towards a modification of the current system in which local grantees would be encouraged to use local evaluations as vehicles for program improvement. Some specific recommendations for improving local Even Start evaluations and for enhancing the ability of Even Start grantees to conduct continuous improvement efforts are given below.

Change the Even Start legislation and preliminary guidance to refer to "local continuous improvement efforts" instead of "local evaluations." Terminology is important, and the past and current use of the term "local evaluation" has become synonymous with a study only of program outcomes, with the implication that local projects will be in trouble if they cannot demonstrate positive outcomes. Use of the term "local continuous improvement effort" would more clearly signal the understanding that Even Start is a difficult program to implement, and the intention that projects are expected to engage in a systematic, ongoing assessment

of the strengths and weaknesses, as well as the outcomes, of their approaches, and to use evaluation data to improve their programs.

Provide guidance to state coordinators and/or local projects on the amount of funding that needs to be spent on a good local evaluation which focuses on program improvement. Without such guidance new projects have no idea of the amount of funds to allocate for this activity, and some do not include a local evaluation line item in their budgets. Suggesting a set amount or a percentage of the total project's budget that should be allocated to evaluation will signal the importance of evaluation and data collection to local program staff.

Help establish a community of local evaluators that can exchange information about useful evaluation approaches by making a mailing list of local evaluators available, by helping local evaluators exchange e-mail addresses, and by having sessions for local evaluators at annual Even Start conferences.

Provide training for local grantees in using data collected at the state and/or national level for continuous improvement at the local level. Local family literacy projects often regard participation in state-level and national-level studies as a burden that offers little or no return. State and national evaluation sponsors could help ensure local use of data by providing training to local staff in (1) accessing data collected for state or national studies, (2) understanding what those data mean at the local level, and (3) using those data to assess the performance level of their projects and to improve unsatisfactory performance.

Provide guidance to local projects on the use of data/evaluation for program improvement. Program improvement currently occurs through trial-and-error, and is based on the hard-learned experiences of program staff. This is one way of improving programs, but greater gains could be made by accompanying these personal, anecdotal methods with a data-based systematic assessment of program strengths and weaknesses. Such systematic continuous improvement efforts have been described by Haslam (1998) in a study of 12 Even Start projects, by Alamprese (1996) in studies of workplace literacy programs, and by Appel (1998) in her work with local Even Start evaluations.

The Department, in the Observational Study of 10 well-implemented Even Start projects, is facilitating and assessing the use of the continuous improvement approach based on the following principles:

Collaboratively set outcome goals for children and families. Local project staff, evaluators, and families enrolled in the program work together to set concrete outcome goals to be achieved.

Devise an intervention to achieve those goals. The family literacy intervention should be designed to achieve the specific outcome goals set in the first step. The intervention should be based on practices shown to be effective in prior research as well as on proven program quality indicators.

Set intervention thresholds necessary to achieve goals. An assumption underlying all family literacy programs is that if a family is to achieve their goals they must participate in the program for a sufficiently long period of time at a sufficiently high level of intensity. Program staff need to define, ahead of time, the minimum intervention threshold that they believe is needed for families to achieve their goals.

Assess progress toward goals with sound measures. Assessing progress on a periodic basis by using sound measures (i.e., measures that have adequate reliability and validity, that have a history of use in similar studies, that are available in appropriate languages, and so on) is the best way of determining whether families have achieved their goals.

Use evaluation to monitor program quality and results, and to target areas for improvement. A comprehensive evaluation will help program staff monitor the quality of their family literacy program. It also will help staff keep track of each family's level and duration of participation in the intervention and assess progress towards goals. Knowing the quality of each program component, the extent to which families have participated, and the degree to which they are making progress will allow program staff to understand why some families do not achieve their goals (e.g., they did not participate at a sufficient level, or program components were of low quality), to identify program components that need improvement, and to target resources accordingly.

Synthesis of Local and State Even Start Evaluations:

Final Report

Since its inception, the Even Start program has undergone evaluative scrutiny at a variety of levels. In addition to the ongoing federally-funded national evaluation, each local project is required by law to conduct an independent local evaluation, and in some states, projects participate in a state-wide evaluation. While the primary purpose of many state and local evaluations is to help improve and focus the program in specific locales, these evaluations may contain information that could be more widely disseminated to program developers and policy makers.

To understand what types of evaluation activities are being undertaken, and to see whether state and local evaluations do, indeed, contain information that would be useful for project improvement and helpful outside of their immediate applications, the U.S. Department of Education (hereafter, the Department) asked Abt Associates Inc. to collect, read, and report on state and local Even Start evaluations. This document is the final report from that project.

This project aimed to gather and review state and local Even Start evaluations, to describe the types of evaluations that were conducted, to describe findings about the outcomes of Even Start programs, and to provide guidance on improving state and local evaluations. In practice, the study focused on local evaluations since few state-level studies were available for review at the time that we were collecting evaluation documents.

1.0 Overview of Even Start

The Even Start Family Literacy Program was first authorized in 1988 as Part B of Chapter 1 of Title I of the Elementary and Secondary Education Act of 1965 (ESEA). The Even Start legislation was amended in July 1991, when Congress passed the National Literacy Act (P.L. 102-73). In 1994, Even Start was reauthorized as Part B of Title I of the ESEA as amended by the Improving America's Schools Act.[3] According to the legislation, the Even Start program is intended to:

> "...help break the cycle of poverty and illiteracy by improving the educational opportunities of the Nation's low-income families by integrating early childhood education, adult literacy or adult basic education, and parenting education into a unified family literacy program...The program shall (1) be implemented through cooperative projects that build on existing community resources to create a new range of services; (2) promote achievement of the National Education Goals; and (3) assist children and adults from low-income families to achieve to challenging State content standards and challenging State student performance standards."
> (P.L. 103-382, Sec. 1201).

Amendments subsequent to the 1994 reauthorization include a requirement, enacted in 1996, that instructional services be intensive. Recent amendments included in the FY 1999 Appropriations Bill for Health and Human Services, Labor, and Education address the need for local evaluations to collect data on program effectiveness and require the Department to provide technical assistance to states and Even Start projects to ensure that local evaluations provide accurate information on the effectiveness of local projects. The legislation also requires states to develop results-based indicators of program quality and to use these indicators to monitor, evaluate, and improve Even Start programs.

Program Operations

To be eligible for Even Start in 1995-96 (or later), a family needed (a) a parent who was eligible for adult education services under the Adult Education Act or who was within the state's compulsory school attendance age range and (b) a

---

3 This description of Even Start refers to the reauthorized law. Projects were not required to implement changes made by that law until program year 1995-96.

child under 8 years of age. Beginning in 1995-96, more teen parents became eligible for Even Start. This new participant group included teen parents – either under or over age 16 – within a state's compulsory school attendance age range, as long as a local educational agency provided for the basic educational services for these parents.[4]

Even Start began as a federally administered program in fiscal 1989, with grants totaling $14.5 million awarded to 76 projects. According to the Even Start statute, when program funding reached $50 million, the program was to be administered primarily by state agencies. This level was exceeded in 1992 when the federal appropriation for 340 projects reached $70 million.

Most Even Start projects now are state administered, and more than $100 million was distributed to some 650 Even Start projects in fiscal 1997. Each state receives funding based on the relative proportion of funds it receives under the Title I allocation formula. States hold grant competitions and make subgrant awards. The statute specifies that each Even Start subgrantee receive a minimum of $75,000 per year, except for one subgrantee per state that may receive less.

Family literacy programs specifically for migrant families, Indian tribes and tribal organizations, and outlying areas are supported through special set-aside funds (5 percent of the total Even Start allocation) and remain under federal administration. These funds support Even Start projects tailored for groups of participants with special circumstances. Since 1993-94, about 10 to 20 each of Migrant Education Even Start (MEES) and tribal Even Start projects have been funded each year.

MEES projects serve a highly mobile population – families moving across several states each year. In addition to economic and educational limitations common to all Even Start families, many migratory parents and children are recent immigrants and have limited English language skills. MEES educational services need to account for great diversity in language and cultural backgrounds of participants. One of the major challenges for MEES projects is to design and deliver meaningful educational services to families who may participate for only a few months before moving. Some MEES projects are designed to follow the same families across states over time; others focus their efforts on families only while they are in their communities.

---

4 Prior to this change regarding eligibility of teen parents, Even Start could only serve families headed by teen parents at least 16 years old or beyond the age of compulsory school attendance *who were not attending school.*

Tribal Even Start projects experience a different set of special challenges. Families tend to be stable geographically but may be widely scattered in remote, rural areas, where families' access to educational and support services provided by other community organizations can be limited. For these reasons, some tribal projects rely on home-based educational services. While such circumstances are common to many rural projects, incorporating materials and activities that are consistent with and promote the Native American heritage is an important objective of many tribal Even Start projects.

In addition to the MEES and tribal projects, discretionary grants for statewide family literacy initiatives and a family literacy project in a prison that houses women and their preschool-aged children are authorized. These projects also are funded and administered directly by the Department.

Design of Even Start Projects

Even Start's premise is that combining adult literacy or adult basic education, parenting education, and early childhood education into a unified family literacy program offers promise for helping to break the intergenerational cycle of poverty and low literacy in the nation. The Even Start program has three related goals:

- to help parents improve their literacy or basic educational skills;
- to help parents become full partners in educating their children; and
- to assist children in reaching their full potential as learners.

The Even Start legislation requires a number of key elements and features to be implemented in all local projects. They include serving families most in need of Even Start services[5]; providing three core services (adult, parenting, and early childhood education) and support services; providing some services to parents and children together and providing some home-based services; integrating educational activities across the three core areas; coordinating service delivery with other local programs; conducting local evaluations; and participating in the national evaluation.

5 The definition of most in need is community-specific and is based on locally established criteria but must be based upon a family's low income and low literacy, as well as other need-related criteria.

Even Start's core services have three components, as specified in the reauthorized legislation:

- ***Adult education and adult literacy:*** high-quality instructional programs[6] to promote adult literacy, including adult basic education (ABE), adult secondary education (ASE), English as a second language (ESL), and preparation for the General Education Development (GED) certificate.

- ***Parenting education:*** high-quality instructional programs to help parents support the educational growth of their children.

- ***Early childhood education:*** developmentally appropriate educational services for children designed to prepare them for success in regular school programs.

Even Start projects also provide support services which are designed to facilitate the provision of core services. Examples of support services are transportation, child care, health care, meals, nutrition assistance, mental health referrals, referrals for employment, counseling, child protective services, referrals for screening or treatment for chemical dependency, referrals for services to battered women, and special care for a disabled family member. Even Start requires support services to be obtained from existing providers whenever possible so that Even Start projects do not duplicate services.

Even Start is intended to benefit families in several domains. Potential outcomes for parents include improved literacy behaviors (e.g., shared literacy events with children and increased reading and writing activities in the home); parenting behavior and skills (e.g., positive parent-child relationships and expectations for children); and educational and employment skills (e.g., improved reading and English language ability and higher education attainment). Goals for Even Start parents also may include growth in personal skills and community involvement. The potential positive impacts of Even Start on children include improved school readiness and achievement (e.g., language development and emergent literacy). Once children enter school, outcomes might include satisfactory

---

6 In April 1996, the Even Start statute was amended to require high-quality, intensive instructional programs. This requirement became effective for projects in program year 1996-97.

school performance, improved school attendance, and a lower incidence of special education and retention in grade.

Variations on the Basic Model

While setting forth major elements required for all Even Start projects, the Even Start legislation allows grantees great flexibility in designing services to meet local needs. The legislation provides Even Start projects with a set of requirements – broad guidelines on what to do. However, decisions regarding how to implement each requirement are left to individual projects. For example, legislation requires high-quality, intensive instructional programs, but projects decide on the frequency and duration of program activities, whether the activities are primarily center-based or home-based, and whether to invent educational curricula from scratch or use a hybrid of existing approaches. Projects decide, based on the availability and quality of local services, which program activities will be supported by Even Start funds and which components will be supported by collaborating agencies.

Most Even Start projects provide, either directly or by working with existing early childhood programs such as Head Start, a center-based early childhood program (Tao, Gamse & Tarr, 1998, p.87). Center-based programs usually incorporate elements of existing curricula designed for young children. Generally, school-age children through age 7 receive Even Start services provided in conjunction with compulsory education activities. Such services may take the form of homework assistance given in before- and after-school child care programs and summer school activities.

Adult education services are provided in a variety of formats by different levels of trained personnel, ranging from volunteers to certified adult education teachers. Some projects offer adult education classes geared toward completing a GED, while others provide general instruction in basic skills such as reading, writing, and math. In some projects, adult education services are focused chiefly on ESL. Projects working with adults who have very low-level basic skills may arrange individual tutoring through organizations such as the Literacy Volunteers of America or provide other types of one-on-one instruction during home visits.

Parenting education is less frequently available through existing community agencies than are adult and early childhood education programs. Thus, many projects rely mostly on Even Start resources to deliver parenting education services. These services may take the form of group discussions, hands-on activities, home visits, and presentations by invited speakers. Topics addressed may include

helping families make use of available learning resources, increasing parents' understanding of normal child development patterns and of their role in their children's education, and training parents on reading to young children.

Educational activities are often offered in institutional settings, e.g., adult education classes in high schools and community colleges, and preschool programs associated with community-based organizations or local education agencies. In some projects, however, particularly those in sparsely populated rural areas, Even Start services may be primarily home-based, with instruction tailored to each family's needs.

2.0 Even Start Evaluation Requirements

Even Start was initiated as a demonstration program. After its first decade, it remains a program that is designed to help demonstrate and learn about the best ways of implementing and delivering family literacy services. To assist with this mission, strong evaluation requirements have been a prominent and continuing feature of the Even Start legislation. Here we briefly discuss Even Start's evaluation requirements at the national, state, and local levels.

National Evaluation

The most recent comprehensive Even Start legislation is the 1994 Improving America's Schools Act, Part B of Title I of the ESEA. That legislation called for:

> "...an independent evaluation of programs assisted under this part--
> (1) to determine the performance and effectiveness of programs assisted under this part; and
> (2) to identify effective Even Start programs assisted under this part that can be duplicated and used in providing technical assistance to Federal, State, and local programs." (P.L. 103-382, Sec.1209).

To carry out this legislative authority, the Department has supported an ongoing national Even Start evaluation in which each local project uses a PC-based management information system to provide the Department with descriptive information on program operations, services offered, participant characteristics, participation rates, and program outcomes; and in which selected projects participate in more intensive studies of program impacts. The national Even Start evaluation is supported with funds reserved from each year's appropriation for the Even Start program.

State Evaluation

The Even Start legislation does not call for state-level evaluation, and Even Start does not provide any funding specifically for state-level evaluations. However, Even Start is administered by state agencies and in the past few years, several states have assumed the responsibility for planning and conducting evaluations of the projects that they have funded. To date, state evaluations generally have been funded either by using state administrative funds, or by local projects combining their evaluation funds to provide a large enough pool to conduct coordinated local

project evaluations that also can serve a satisfactory state-level study purpose. Though our current understanding of the nature of state-level studies is limited, we do know, for example, that some states are putting their own management information systems in place, others are conducting pre-post evaluations in each funded project, and still others are working with grantees to devise a mutually-agreeable common set of measures.

When we inquired about the availability of evaluations, few states had reports available for public release. However, many state coordinators said that they either were planning evaluations or were in the process of conducting them. These state-level studies typically were being performed through contracts with independent contractors.

Recent amendments to the Even Start program in 1998 include a requirement that states develop results-based indicators of program quality and to use these indicators to monitor, evaluate, and improve Even Start programs. The Department will be able to work with states in developing their indicators of quality through the Statewide Family Literacy Initiative grants funded through the same set of amendments to Even Start.

Local Evaluation

In addition to language requiring a national evaluation, the most recent Even Start legislation contains the following language about local evaluations:

> "Each program assisted under this part shall...provide for an independent evaluation of the program." (P.L. 103-382, Sec. 1205).

When Even Start was administered at the federal level (from 1989 to 1991), it was recommended that local projects allocate $5,000 to $10,000 per year to cooperate with the national evaluation and to conduct an independent local evaluation. This rule of thumb for the amount to be spent on a local evaluation has been maintained in some states, while other states offer little or no guidance on the amount of funds to be spent on a local evaluation.

Assuming an average expenditure of $5,000 per project per year, the total amount of federal funds spent on local evaluation activities has grown from about $380,000 in 1989 when there were 76 Even Start projects to more than $3.25 million in 1997 with about 650 projects. This is a substantially larger amount than what has

been spent on the national evaluation during the past decade (about $500,000 to $750,000 per year).

While the Even Start legislation does not provide local projects with much direction on how to spend their local evaluation funds, the Department has provided several forms of guidance about local evaluations to Even Start project directors. Further, since Even Start is now a state-administered program, each state can provide its own advice to local projects on how to best comply with the local evaluation requirements.

As part of the first national evaluation (1989-90 through 1992-93), the Department asked the national evaluation contractor to work with local projects so that they would be able to submit evidence of effectiveness to the Department's Program Effectiveness Panel (PEP). Consequently, sessions on preparing PEP submissions were held at national evaluation meetings and interested projects were trained in how to use national and/or local data to make a case for the effectiveness of their projects. In response to questions raised by local project directors at national evaluation meetings, Department staff responded that while local evaluation monies were intended to fund a local study that would complement the information collected for the national evaluation, local evaluation funds also could be used to pay for various tasks associated with complying with the national evaluation, e.g., interviewing parents, entering data into the computerized MIS, and attending national evaluation conferences. In these early days of Even Start, many local projects were uncertain about the purpose and requirements of their local evaluation.

As part of the second national evaluation (1993-94 through 1996-97) the topic of how to conduct local evaluations continued to be raised and discussed at national evaluation conferences. For example, the June 1994 Even Start Evaluation Conference contained a session titled "Two paths to program validation: PEP and the National Family Literacy Dissemination Project." This built on the emphasis seen in the first national evaluation on having local projects prepare and submit evidence of their effectiveness to the Department in order to possibly obtain additional funding as a validated approach.

Another session at the 1994 meeting was titled "Designing and implementing a local evaluation." At this session, staff from RMC Research Corporation worked with participants to design a "vision for local evaluation guidance" that formed the

basis of a Local Evaluation Guide for Even Start projects (Dwyer, 1998).[7] The vision statement specified that projects were to conduct "comprehensive local evaluations," that among other things assess "effectiveness in terms of participant outcomes," and that are "complementary to the national ESIS reporting system" (Exhibit 1).[8]

In the third national evaluation (1997-98 through 2000-01) there has been a continued emphasis on the use of national evaluation data for local purposes, as evidenced by the annual preparation of Project Profiles which compare local descriptive and outcome statistics to data from state and national aggregates. The emphasis on gathering information for PEP submissions disappeared[9], replaced in part by an Observational Study of Even Start projects that provided evidence of positive outcomes over at least a two-year period.

The most recent Even Start guidance provided by the Department contains the following language:

> "Each Even Start project is required to provide for an independent evaluation of the project. See section 1205 (10) of the Act. These evaluations provide local projects, States, the Department, and the Congress with objective data about the activities and services provided by the project, the participants served, the retention rates of those participants, and the success of the families in the project." (Department of Education preliminary guidance).

The Department interprets the requirement for an independent local evaluation to mean that the evaluation must be done by a person not directly involved in the day-to-day administration of the project. Thus, the evaluation could be done by an evaluator who is external to the LEA, or by a person within the LEA who does not work directly for the Even Start project.

Other related resources available to local projects include a *Guide to Quality for Even Start Family Literacy Programs* and a *Program Self-Assessment*. Prepared for the Department of Education by RMC Research Corporation, the Guide provides

---

7 An incomplete draft of the guide was distributed during a series of evaluation conferences in March and April 1998 for use by local projects.

8 The ESIS was the second iteration of PC-based Even Start information systems. It was designed to collect information on the nature and intensity of all Even Start projects and on the characteristics and participation patterns of all participating families.

9 The National Diffusion Network and Program Effectiveness Panel lost their funding during this period.

projects with a discussion of the components of an Even Start project and ways to ensure the quality of those components. Projects are supplied with a self-assessment measure that they can use to rate the quality of their projects. While projects are not asked to report on the results of the self-assessment in their independent evaluation reports, we have heard that projects do use the Guide and Self-Assessment in analyses of their programs.

EXHIBIT 1
"VISION FOR LOCAL EVALUATION GUIDANCE"
(QUOTED FROM JUNE 1994 EVALUATION CONFERENCE MATERIALS)

"The Department of Education expects all Even Start projects to put into place soon comprehensive local evaluations that yield information useful for the ongoing management of programs, provide the basis for program improvements, and meet accountability demands of funders and collaborators. For programs that have attained operational stability, local evaluations should also address assessment of effectiveness in terms of participant outcomes, including those outcomes that can be compared with national results.

The Department of Education requires Even Start projects to conduct local evaluations in addition to being active participants in the national evaluation of the program. Local evaluation data collection should be complementary to the national ESIS reporting system, taking advantage of required national data for local uses. Careful planning of an information system is a key part of local evaluation design to facilitate maximum use of the national evaluation tools and database results.

Because all projects are not currently required to provide outcome data, it is important for the future development of Even Start to encourage, document, and share what is known about an array of program effects. Local evaluations offer the opportunity to experiment with measurement of the wide array of short, intermediate, and long-term outcomes that result from Even Start programs. Information about participant outcomes measured in a variety of ways has great importance for depicting the value of Even Start interventions to decision makers. The local data is needed to augment what is collected via ESIS.

For projects that yield exemplary outcomes, there are two national avenues of dissemination, the Program Effectiveness Panel and the National Family Literacy Project. Approval by either dissemination mechanism requires systematic collection of implementation and outcome data.

For all the above reasons, planning the design of a local Even Start evaluation involves numerous choices and a high degree of communication between project personnel and evaluation staff. Choices made early on in the planning will affect the ways in which comparisons can be made later to gain both power and efficiency for local efforts. Anticipation of eventual uses for information pays off later.

To encourage thoughtful planning, the Department of Education (with the help of experienced Even Start evaluators) is preparing a guide for designing local Even Start evaluations. The guide will function both as a planning tool for local project directors to use in dialogue with potential program evaluators during program start-up and also provide guidance for evaluators during the design phase. The types of evaluation questions and outcomes posed by the guide, the suggested linkages to the national evaluation, and the possible sources of relative comparisons will provide the basics for taking advantage of local project data.

We anticipate that SEA monitors will also find the guide helpful in their support of local project development and improvement. Further, the guide will help SEAs use the local project data they will receive as part of the ESIS system."

3.0 Study Methods

Research Questions

At the start of this study, the Department posed a series of questions about what local evaluations look like and how they are conducted, as well as questions about the kinds of findings reported in local evaluations. Analyses of local evaluations are limited in the extent to which they can contribute to our knowledge about Even Start's program operations and effectiveness. This is due to several characteristics of local evaluations, such as: variation in the kinds of studies done at the local level, limitations of local evaluation designs, and issues with the quality of local evaluations. This report describes local evaluation studies in Section 4 and summarizes a limited set of the findings from those studies in Section 5.

Obtaining Evaluation Reports

During 1996-97 there were about 650 local Even Start projects, and the law specifies that each project ought to conduct a local evaluation. At the same time, some states have planned and are conducting their own evaluations. For the present study, we hoped to collect at least 100 local and state evaluation reports, though we had no idea how many reports could actually be obtained.

Local evaluations are conducted for the benefit of local projects, and those projects are not required to transmit local evaluation reports to the Department. However, some State Coordinators require their grantees to submit copies of their local evaluations. Therefore, we sent a letter to all State Coordinators, describing the planned synthesis and asking for their help in obtaining copies of state and local studies.

Obtaining local evaluations was difficult (see Exhibit 2). A few State Coordinators sent us copies of the local evaluations they had received. In other cases the State Coordinator did not have copies of local evaluation reports, but they were able to send us a list of local projects or local evaluators and suggested that we call projects directly in order to obtain evaluation reports. In 33 cases, the State Coordinator did not respond to the request.

With limited time and funds, we ended up with a sample of convenience – a total of 4 state evaluations (1 each from Hawaii and Nevada, and 2 from Oregon) and 118 local evaluations from 19 states. We do not know how well the 118 local evaluations represent all local evaluations.

Receiving and Logging-In Reports

As we received evaluation reports we created a data base by logging-in each report and recording selected pieces of information. The resulting data base allowed us to produce a complete listing of all reports received in this project and to describe the universe of state and local reports submitted to us.

The log-in form was developed on the basis of (1) a review of the evaluation literature concerned with assessing the quality of research studies, and (2) the needs of this project to capture a variety of study types. There is substantial literature on methods of assessing the quality of an evaluation (e.g., Cooper & Hedges, 1994). However, this literature focuses almost completely on the quality of evaluations which seek to determine the effectiveness of a given program, and hence is concerned with issues of random assignment, control groups, sample sizes, and so on. The same holds for the very large literature on meta-analysis. We examined samples of the coding schemes used in these studies and used variables which were most relevant to the present context.

We found little research literature on assessing the quality of implementation studies. This is unfortunate because we expected that many of the studies conducted by Even Start projects and states would focus on implementation issues. Because we did not want to neglect evaluations that focused on implementation issues, we devised our own coding scheme for implementation studies.

Describing Methods Used to Conduct Local Evaluations

We used data from the log-in form on all 122 evaluations received to describe the kinds of studies that were done at the local level. These data allowed us to calculate the percentage of studies that focused on program outcomes and to describe the design of those studies, the time period covered by the studies, the sample(s) measured, and the kinds of measurements that were made. We also calculated the percentage of studies that focused on implementation issues to describe the design of those studies, the sample(s) measured in those studies, and the kinds of measures that were used.

Describing Findings from a Subset of Evaluations

We selected a subset of 24 local evaluation reports that focused on Even Start's outcomes, read those reports carefully, and used the data and findings from those reports in the descriptions in Section 5. The reports included 3 state reports with outcome data, 3 local reports based on a non-equivalent comparison group design, and 18 reports based on a one-group pre-post design with reasonably large samples and measurement of both adults and children.

Each of the 24 reports addressed the question of program outcomes, and we had to summarize data reported in differing levels of detail from studies that used differing (if overlapping) measures, varied data collection strategies and schedules, and different analytic strategies. In recent years, this type of challenge has increasingly been addressed through formal meta-analysis of cross-study data on program services, activities and implementation, and outcomes. However, coding studies for a meta-analysis is a time-consuming and expensive process which was beyond the resources of this project.

Instead, we prepared a narrative synthesis of findings based on a series of tables created to summarize the salient features of each study, as well as the outcome findings for each study. This strategy is often used in small-scale reviews. The general approach was to create tables for different research questions in which each of the 24 reports is a row and the cells of the table contain summary statements and/or summary data from the studies about the research question. The discussion in Section 5 of this report was drawn from reviewing the evidence contained in these tables.

EXHIBIT 2

NUMBER OF EVALUATIONS OBTAINED, BY STATE (TOTAL N=122)[10]

STATE	N OF LOCAL (STATE) EVALUATIONS OBTAINED	STATE	N OF LOCAL (STATE) EVALUATIONS OBTAINED
Alabama	0	Montana	0
Alaska	5	Nebraska	0
Arizona	2	Nevada	0 (1 state)
Arkansas	0	New Hampshire	0
California	3	New Jersey	0
Colorado	1	New Mexico	0
Connecticut	0	New York	0
Delaware	0	North Carolina	0
Dist. of Columbia	0	North Dakota	6
Florida	1	Ohio	17
Georgia	0	Oklahoma	8
Hawaii	1 (1 state)	Oregon	0 (2 state)
Idaho	2	Pennsylvania	18
Illinois	10	Puerto Rico	0
Indiana	0	Rhode Island	0
Iowa	0	South Carolina	0
Kansas	0	South Dakota	0
Kentucky	4	Tennessee	17
Louisiana	0	Texas	1
Maine	0	Utah	0
Maryland	0	Vermont	0
Massachusetts	0	Virginia	0
Michigan	6	Washington	0
Minnesota	3	West Virginia	0
Mississippi	7	Wisconsin	0
Missouri	1	Wyoming	5

10 We received an additional 55 local evaluation reports after the date when we were able to read them for the review. This includes 28 reports from California, 24 from New York, 2 from Tennessee, and 1 from Indiana.

4.0 Description of Local and State Evaluations

Little is known about the characteristics of state and local Even Start evaluations[11]. This is the first study to describe some of the types of evaluative activities that are being undertaken with Even Start funds at the state and local levels. Research questions in this area include the following:

- What are the purposes of local evaluations? To what extent are local evaluations used to improve program practices?
- What is the quality of state and local Even Start evaluations? Who conducts these studies?
- What are the characteristics of state and local Even Start evaluations? Do they focus on implementation? On outcomes?

Using Local and State Evaluations for Program Improvement

Local evaluations are conducted for multiple purposes. First and foremost, Even Start projects have to comply with the legislative requirement for a local evaluation. Given this requirement, many local evaluations choose to focus on project outcomes in an attempt to provide evidence so that local project directors can obtain political and/or financial support for the program from school boards, civic organizations, corporate sponsors, or foundations.

Ideally, Even Start projects would collect and use data as part of an ongoing continuous improvement effort and the local evaluation activity would be conceived and implemented with that in mind. The evidence contained in the reports that we reviewed shows that Even Start projects rarely engage in the systematic use of data to manage and improve their programs. Instead, program improvements/alterations typically are made on the basis of anecdotal evidence obtained through observations and stories gathered from the personal experiences of program implementers. A few reports that we reviewed noted recommendations from the previous year, described whether they had been addressed, and provided additional recommendations for the current year.

---

11 For this discussion we combine local and state evaluations, since there are only 4 state evaluations in our sample.

One reason for the apparent lack of use of data to improve Even Start projects is the distinction between the work that is done for a local evaluation and the work that is documented in the local evaluation report. Based on discussions with Even Start project directors and evaluators it appears that local evaluations may focus more on program improvement than would appear from reading the reports from those evaluations. When faced with limited time and resources, a local evaluator is likely to prepare a report which documents the gains made by families, rather than the types of programmatic improvements that he/she might have diagnosed and recommended, and the response that was made by the project.

Another reason is a misunderstanding of the purpose of the local evaluation, and who can benefit from the data collected. Perhaps local evaluation reports rarely provide information on continuous improvement efforts, testing for placement, and other local uses of data because Even Start project directors and/or local evaluators do not see these as important aspects of a local evaluation. The emphasis on simply reporting outcomes in local evaluations is understandable since the Even Start legislation and the guidance that has been provided to grantees both refer to studies of effectiveness and outcomes. Ideally, local evaluators would work with local projects to help them choose appropriate assessments and evaluation techniques to measure progress toward program goals. Even more important is the need for local evaluators to help local projects with interpretation of data and use of what is learned to improve services.

Finally, an important gap exists between the data and the conclusions of many local evaluation reports. While local evaluations almost always report glowingly positive conclusions about the effectiveness of Even Start projects, an independent reading of the same reports would rarely come to the same conclusions. These conflicting interpretations of the same data occur for two reasons. First, as described in this report, data gathered in local evaluations are almost always collected only on participants in the program. Those data typically show that children and adults improve over time on relevant outcome measures, and local evaluators generally report that the program helps participants. However, attributing improvements in test scores or other outcomes to Even Start is incorrect without considering alternative explanations. For example, children and adults make gains on measures such as the PPVT, the PSI, the TABE, and the CASAS due to normal development and maturation. Without a control group to assess the size of the normal developmental gain, it is inappropriate to attribute these gains to Even Start.

A second reason for the discrepancy between the conclusions reached by local evaluators and unbiased readers is that local evaluators are rarely

independent evaluators. Although local evaluators are not allowed to be Even Start staff members, sometimes they are school district employees (e.g. a curriculum specialist or a school principal), sometimes they are self-employed researchers with small practices, and sometimes they are university-based consultants with varied research training. Local evaluations are rarely done by independent research firms. In any case, continued funding often rests on a local evaluator's ability to make some positive conclusions with whatever data are collected.

These evaluation issues have implications for using data for continuous improvement. If it is politically unacceptable for an evaluation to point out program weaknesses, to state that a program is not meeting its goals, or to demonstrate that children or parents are not attaining hoped-for literacy skills, then it is difficult to conclude that a program needs improvement.

Quality of Local and State Evaluations

The great majority of local evaluations are conducted by persons outside the grantee agency. The typical local evaluator is a university-based consultant, though many local evaluations are done by research organizations, and some are done by an outsider in concert with a grantee staff member. Many local evaluators have advanced degrees in a relevant field while others seem to have fewer formal credentials. While we have no direct evidence about the training of local evaluators, there is great diversity in the quality of local studies, just as there is great diversity in every aspect of Even Start.

Some local evaluation reports are clearly conceptualized, carefully conducted, and well written. They often provide information describing Even Start families, the amount and level of participation in Even Start, the gains made by Even Start participants on various measures, and the degree to which parents are satisfied with the program. Some of these studies also provide information about the ways in which local collaborations are or are not working, methods of family recruitment that seem to work well, and recommendations for improving project operations. These well-done studies have the potential to affect positively the development of local projects. If these studies are being done for $5,000 to $10,000 per year, local projects are getting a lot for their money.

Other local evaluations are of quite poor quality – incoherently conceived and presented, poorly written, and of little assistance to a local project. They are evaluations in name only, conducted to satisfy the letter of the law. Anecdotal information from local evaluators suggests that variation in state-level policies about

the importance of local evaluation and the amount of funding to be spent on local evaluation contributes to the variation in the quality of local studies; high-quality local evaluations cannot be purchased for under $2,000 per year.

Description of Implementation Studies

Of the 122 studies in the sample, we judged that 113 (93%) contained data about the implementation of the program being studied (Exhibit 3). This matches reports from the Even Start Information System in which 94% of all Even Start projects reported that they collected data on participants, services, and interagency collaboration, and 95% reported that they assessed program implementation.

Design of implementation studies. Almost all of the local implementation studies (95%) included some sort of description of the project structure and activities undertaken by families served by Even Start. This ranges from one-paragraph statements about the nature of the project and services provided to detailed, multi-page charts showing the exact length and nature of the planned services in each component, as offered or delivered by several different service providers.

Almost two-thirds (62%) of the local implementation studies provided information about the level of participation in Even Start. Typically, this includes data on the average length of time (months) that families participate; less often it includes data on the average amount of time (hours per month) that families participate. Many studies commented on the difficulty of recruiting and retaining families in Even Start, but none contained a systematic accounting of the number of families contacted, the number that showed up for at least one contact, the number that officially enrolled in Even Start, and the numbers that stayed in the program for different lengths of time.

Sample for implementation studies. Data to understand local implementation issues most often came from the local project director (81%), and from studies which intended to measure all Even Start teachers (72%) or all Even Start parents (70%). Rarely did local evaluations call for samples of teachers or parents. This is a reasonable approach since for small studies it generally is easier to collect data from all potential respondents than it is to select a sample and collect data only from the sample; for Even Start projects, the universe of respondents often is small enough to warrant a census approach.

Measurement methods used in implementation studies. Self-report was the most common measurement approach for Even Start local implementation

studies. This strategy was used to obtain information from project directors (75%), teachers (72%), and parents (40%). Other commonly used measurement methods included abstraction of information from project records (66%), logs of activities or participation (32%), observation (23%). Only 15% of the local implementation studies relied on data collected for the national evaluation.

Description of Outcome Studies

Of the 122 studies in the sample, we judged that 94 (77%) contained some information on program outcomes (Exhibit 4). This is less than the 93% of all Even Start projects which reported through the Even Start Information System that they assessed growth in child and adult literacy, and parenting skills.

Design of outcome studies. Many local evaluations reported on program outcomes. Without the resources to conduct high-quality outcome studies, local evaluators often measured the gains which families in the program made on standardized tests, but were not able to determine whether those gains are larger or smaller than would be expected in the absence of the program.

The most common design for local evaluations was the one-group pre-post study, used in 76% of the outcome studies. In this design, Even Start families are assessed as they enter the program and again at a later point in time, often at the end of a school year, or when they leave the program. No control or comparison group families are measured in this design, meaning that while the local evaluator can calculate the gains made by Even Start families, there is no way of knowing how much the families would have gained if they were not in Even Start.

While the lack of a comparison group makes the one-group pre-post study a weak design for estimating the effectiveness of Even Start, pre-post data can be used to assess the effectiveness of a project at helping families to meet a pre-identified set of literacy standards (e.g., at entry to kindergarten, 80% of the children who participated in Even Start will be able to perform tasks a, b, and c), and to determine whether parents and children achieve above, at the same level, or below the levels of parents and children in national norms groups on tests of literacy skills. Though it would be useful to program staff, this type of analysis is rarely included in local evaluations, in large part due to the difficulty of agreeing on performance standards to be met by program participants. The indicators that states must develop under recent amendments to Even Start have the potential to address this difficulty.

In a weaker design, 31% of the outcome studies used a one-group post-only design in which Even Start families are administered a posttest, but not a pretest. This design allows calculation of whether Even Start adults and children achieve at a given level, but it is not possible to determine how much was gained. Again, it would be possible to use data from this design to assess the performance of Even Start participants against a set of literacy standards, but setting those standards is a task which is only now beginning with the state work on indicators of program quality.

Only 10% of the local outcome evaluations used a two-group quasi-experimental design in which the gains of Even Start families were compared to the gains of families in a non-equivalent comparison group, e.g., families in a parallel program, children in Head Start, etc. Local outcome evaluations rarely have the resources to conduct a study in which funds are used to collect data from non-Even Start families.

Finally, no local outcome evaluation used an experimental design in which families eligible for Even Start are randomly assigned to participate in the program or in a control group. While this is the strongest approach for estimating the effectiveness of Even Start, its non-use in local settings is not surprising, given that designing and implementing a randomized study is a costly, time-consuming enterprise that requires considerable expertise. Local evaluations, rarely if ever funded for more than $10,000 per year, cannot be expected to undertake this type of expensive and complicated approach.

Time period covered by outcome studies. Families were followed for one project year in 75% of the local outcome studies, and for more than one year in 11% of the outcome studies. In general, these multi-year studies tracked children into the public schools in an attempt to learn about school-based child performance. The focus on performance over a one-year period is reasonable because most families participate in Even Start for a year or less.

Measurement methods used in outcome studies. Local outcome evaluations used many measurement methods. Children were most often measured by administering a test such as the PreSchool Inventory (PSI), the Peabody Picture Vocabulary Test (PPVT), or the PreSchool Language Scale (PLS) (64%); through parent interviews about the child's behaviors or progress (38%); or through teacher reports (29%). Adults were most often measured by administering the Comprehensive Adult Student Assessment System (CASAS) or the Tests of Adult Basic Education (TABE) (70%); through self-reports (65%); or through teacher reports (27%). Some of these measures (the PPVT, PLS, and TABE) have national

norms which can be used as one basis of comparison for gains made by Even Start adults and children. The PSI has its own Even Start "norms" based on developmental data collected in the first national evaluation.

Other forms of measurement included abstraction of data from school or project records (35%), observation of the child/adult (19%), observation of the family/home (11%), and use of data from the national evaluation (9%). Given its high cost, it is not surprising that data collection through observation is seldomly used. The same holds for teacher reports; although it is relatively easy for teachers to complete a rating scale for a given child, it requires substantial resources to "track" Even Start children into many different public schools and negotiate to obtain the time of teachers to do "non-school" work.

EXHIBIT 3
IMPLEMENTATION STUDIES:
DESCRIPTION OF STATE AND LOCAL EVALUATIONS

VARIABLE	PERCENTAGE (N) (TOTAL N = 113)
Study Design	
Project structure/description/activities	95% (107)
Level of participation	62% (70)
Sample	
Project director	81% (91)
Teachers (all)	72% (81)
(random/representative sample)	2% (2)
(convenience sample)	1% (1)
Parents (all)	70% (79)
(random/representative sample)	0% (0)
(convenience sample)	4% (5)
Measurement Methods	
Project director report	75% (85)
Teacher report	72% (81)
Abstraction from project records	66% (75)
Parent report	40% (45)
Log of activities/level of participation	32% (37)
Observation	23% (26)
Report from ESIS data	15% (17)

EXHIBIT 4
OUTCOME STUDIES:
DESCRIPTION OF STATE AND LOCAL EVALUATIONS

VARIABLE	PERCENTAGE (N) (TOTAL N = 94)
Study Design	
Randomized experiment	0% (0)
Quasi-experiment (two groups)	10% (9)
One-group pre-post study	76% (71)
All project families	60% (56)
Random/representative sample of families	0% (0)
Convenience sample of project families	6% (6)
One-group post only study	31% (29)
All project families	18% (17)
Random/representative sample of families	0% (0)
Convenience sample of project families	1% (1)
Comparison to norms	12% (11)
Follow-up of project participants	6% (6)
Cross-sectional change only	5% (5)
Time Period Covered by Study	
Families followed for one project year	75% (70)
Families followed for more than one year	11% (10)
Measurement Method	
Adult direct assessment	70% (66)
Adult report on adult	65% (61)
Child direct assessment	64% (60)
Parent report of satisfaction	48% (45)
Adult report on family/home	42% (39)
Adult report on child	38% (36)
Record abstraction	35% (33)
Teacher report on child	29% (27)
Teacher report on adult	27% (25)
Observation of child and/or adult	19% (18)
Observation of family/home	11% (10)
Report from ESIS data	9% (8)

5.0 Evaluation Review

This section summarizes findings about Even Start's outcomes based on the 24 studies that were reviewed in detail. Rather than using the interpretation or conclusions reached in the individual evaluation reports, we reviewed the data/findings presented in the reports and draw our own conclusions.

Due to the great variation in questions addressed, methods used, measurements made, and approaches taken in local evaluations, the information contained in the evaluation reports we reviewed does not allow us to draw a consistent set of conclusions about the implementation or the effectiveness of Even Start projects. This is not a criticism of local evaluations. Rather, it reflects the substantial variation in the purpose and nature of local evaluations. Hence, in this section we focus on the range of outcomes described by local projects and, where possible, provide national comparisons.

Outcomes for Children

Cognitive outcomes. Local evaluations showed a consistent pattern of positive gains on several different measures of child development including the PSI, PLS-3, PPVT, pre-IPT, and COR. Each of these measures was administered on a pre-post basis by three or more projects, and in almost every case, positive gains were observed. When significance testing was done, the gains generally were statistically significant.

The size of the gains is consistent with the size of the gains observed in the national Even Start evaluation and in other family literacy studies which administered these same measures. For example, annual PSI gains in the local evaluations were on the order of 5 to 11 raw score points, roughly the same as the 7.3 points per school year gained by children in the NEIS portion of the first national Even Start evaluation,[12] the 5.1 points per school year gained by children in the Sample Study portion of the second national evaluation,[13] and the 4.5 points per

---

12 Even Start children gained an average of .91 points per month on the PSI as measured by the first national evaluation's NEIS data collection (St.Pierre, et al, 1995, p. 168). School year gains are estimated to be 7.3 points in an 8 month school year (.91 points per month * 8 months = 7.3 points).

13 Even Start children in the second national evaluation's Sample Study gained an average of .64 points per month on the PSI (Tao, Gamse & Tarr, 1998). School year gains are estimated to be

school year gained by children in the Even Start In-Depth Study[14]. Similarly, the annual PPVT gains in the local evaluations were on the order of 5 to 12 points, roughly the same magnitude as the 7.5 points gained by children in the NEIS portion of the first National Even Start evaluation[15] and somewhat greater than the 3.2 points per school year gained by children in the Even Start In-Depth Study[16].

Parent ratings of effects on child. Parents were asked questions about what their child learned while in Even Start, whether their child enjoyed Even Start, and whether their child was helped by Even Start in a variety of areas, e.g., more successful in school. Parents reported, among other things, that Even Start children benefitted from the program, learned to get along better with others, enjoyed reading more, and were more likely to succeed in school.

Developmental screening. Projects administered many developmental screening instruments including the Denver Developmental Screening Inventory, the Batelle, the Early Screening Inventory, the DIAL-R, the Brigance, the Quality Reading Inventory, the HOSTS survey of Developmental Tasks, the Normal Developmental Checklist, the Learning Accomplishments Profile, the Kindergarten Readiness Skills Checklist, the Iowa Tests of Basic Skills, the Psychomotor Skills Inventory, the Early Childhood Assessment, and several locally-developed measures. Even Start children were generally seen as scoring within age-appropriate norms on standardized tests or as making improvements or gains.

Teacher ratings. Preschool and elementary school teachers were asked to evaluate the school readiness behaviors and skills of children using the Teacher Assessment of Achievement, the Growing Up Strong Observable Behavior Checklist, as well as several locally-developed measures. Teacher ratings of Even Start children's behaviors, skills, and interactions typically increased as those

---

5.1 points in an 8 month school year (.64 points per month * 8 months = 5.1 points).

14 Even Start children in the first national evaluation's In-Depth Study gained an average of 10.2 points on the PSI over an 18 month period (St.Pierre, et al, 1995, p. 162). School year gains are estimated to be 4.5 points in an 8 month school year ((10.2 points/18 months) * 8 months).

15 Even Start children gained an average of .94 points per month on the PPVT as measured by the first national evaluation's NEIS data collection (St.Pierre, et al, 1995, p. 173). School year gains are estimated to be 7.5 points in an 8 month school year (.94 points per month * 8 months = 7.5 points).

16 Even Start children in the first national evaluation's In-Depth Study gained an average of 7.1 points on the PPVT over an 18 month period (St.Pierre, et al, 1995, p. 171). School year gains are estimated to be 3.2 points in an 8 month school year ((7.1 points/18 months) * 8 months).

children aged; teachers (kindergarten and later grades) were likely to rate most Even Start children as developmentally ready for school or as doing well in school.

Observations were done of parent/child reading, child verbal interactions, and language sophistication in interactions with peers. There was no change over time in parent/child reading, but observations did reveal positive changes in communication and interaction with peers.

School record abstraction of attendance and grades showed that Even Start children have higher attendance rates than the general school population; analysis of school grades was inconclusive.

This review of the data provided by local evaluations shows that children in Even Start made pre-post gains on several different measures of cognitive development, consistent with the gains observed in the national Even Start evaluations. In addition, regardless of the particular measure used, local evaluation reports invariably reach positive conclusions about the effectiveness of Even Start on participating children.

Outcomes for Adults

Cognitive outcomes. As was the case for child-level outcomes, local evaluations showed a consistent pattern of positive gains on several different measures of adult development including the CASAS, the IPT, and the TABE. Statistically significant gains were observed in almost all projects that administered these measures. In addition, most projects that collected data on GED status reported that substantial numbers or percentages of adults attained a GED during their participation in Even Start.

The local evaluations that we examined report annual CASAS gains ranging from -5 points to +9 points, certainly in the same range as the average 4.6 point per year gain for adults in the NEIS portion of the first national Even Start evaluation,[17] the 4.2 point per year gain for adults in the Sample Study portion of the second national Even Start evaluation[18], and the 1.6 point per year gain for adults in the

---

17 Even Start adults gained an average of 4.6 points in a school year on the CASAS reading scale as measured by the first national evaluation's NEIS data collection (St.Pierre, et al, 1995, p. 189).

18 Even Start adults who participated in the second national evaluation's Sample Study gained an average of 4.2 points in a school year on the CASAS reading scale (Tao, Gamse & Tarr, 1998, p.145).

Even Start In-Depth Study[19]. The CASAS has a standard deviation of 16 points, so the roughly 4 point gains that have been seen in many studies correspond to about one-quarter of a standard deviation, a small difference by social science standards.

TABE scores were generally reported by local evaluations in grade equivalent units, with gains ranging from 0.2 grades per year to more than 2.0 grades per year, depending on the project and the subtest. As is the case with other parent measures, these are in the same range as the average gains seen in the national Even Start evaluation (about 1.3 grades per year in reading and 0.9 grades per year in math).[20]

In addition to the CASAS, IPT, and TABE described above, parents were administered various tests including the Gates-MacGinitie Reading Inventory, and the Basic English Skills Test, the Language Assessment Scale. Pre-post results on these assessments were mixed, with some studies finding positive gains while others saw no growth.

Parent surveys of personal growth and attitudes toward and involvement in education. Parents were asked to report their perceptions about the effects that Even Start had, e.g., whether they learned life skills, made friends, improved math and reading skills, improved self-esteem, improved job skills, learned time management skills, liked school more, were more involved in their child's school, volunteered more in their child's school, and learned workforce readiness skills. Parents also were asked to complete rating scales to assess the effects of Even Start in areas such as social support, family stress, self-esteem, and depression. Parents almost always had very positive perceptions about Even Start's effects on their lives and the lives of their children. However, there were few, if any, significant changes over time on rating scales of social support, family stress, and depression.

Parent self-reported educational level and employment status. Many projects asked parents (or otherwise kept track) about their educational level and employment status. Most projects that reported on these outcomes did not provide data adequate to assess change over time, e.g., projects reported the number of

---

19 Even Start adults who participated in the In-Depth Study gained an average of 3.7 points over an 18-month period on the CASAS (St.Pierre, et al., 1995, p. 186). School year gains are estimated to be 1.6 points in an 8 month school year ((3.7 points/18 months) * 8 months).

20 Even Start adults who participated in the second national evaluation's Sample Study gained an average of 1.3 grade equivalents per year on the TABE reading scale and 0.9 grade equivalents per year on the TABE math scale (Tao, Gamse & Tarr, 1998, p. 149)

parents who obtained a job or an educational credential but did not report the total number of participating parents, or the number who were looking for work.

Parent surveys about goals. Parents were asked about the goals that they established at entry to Even Start and whether those goals were achieved. Parents reported that they were very successful at meeting the goals that they had set at entry to Even Start.

Parenting measures. Across the 24 projects, parents were asked to complete the Home Screening Questionnaire, the Bowdoin Inventory of Parent Beliefs and Practices, and the Systematic Training for Effective Parenting scales, in addition to many different locally-developed questionnaires about knowledge of parenting practices, parenting responsibility, and reading behaviors. Parents also were asked to report their perceptions about the effects that Even Start had in areas such as reading books to their child, talking to their child about school, taking their child to the library, becoming a better parent, knowing about child development, communicating better with their child, and helping their child with homework. Local evaluations generally were unable to detect significant changes when using rating scales of parent behaviors and knowledge. However, parents almost always had very positive perceptions about Even Start's effects on parent/child interactions.

Observations of parenting behaviors were conducted during home visits and during parent and child together (PACT) time. Projects that conducted this type of data collection reported positive changes in parenting behaviors.

Teacher ratings. Teachers were asked to rate parents in areas such as parent involvement in school, communication with school, and workforce readiness skills. There were no pre-post changes in teacher ratings in these areas.

These findings from local evaluations match findings from prior research which show generalized parent satisfaction with almost any social program (McNaughton, 1994) as well as small amounts of cognitive growth, but few measurable changes over time in social-emotional development, or parenting skills as measured by rating scales.

Impacts on Children and Adults

A program "impact" is different from a program "outcome." The outcome of a program is the gain or growth or change that is observed in participants during their time in the program. For example, the statement that "children in Even Start gain "x"

points on the PSI during a program year" describes a program outcome. The impact of a program is the difference between the gain observed for program participants (the outcome for people in the program) and the gain observed for nonparticipants (the outcome for people who were not in the program). For example, the statement that "children in Even Start gained "y" points more than children not in Even Start" describes a program impact.

Impacts compare the outcomes for Even Start participants with some standard, for example, the outcomes achieved by a control or comparison group, or national norms, or the gains normally achieved by Even Start participants. While each of the 24 reports that we reviewed provided information on outcomes, only half of the reports tried to gauge the size of these outcomes by comparing them to the outcomes observed in other studies. Those that did, relied on a range of comparisons.

- ***Comparison to national data.*** Six of the 24 studies compared local Even Start gains on the PSI, PLS-3, CASAS, and HSQ to average gains reported in the national Even Start evaluation. As would be expected, the gains in the 6 studies were larger than national gains in some cases and smaller in others.

- ***Comparison to national statistics.*** One project found that when compared to data presented in the National Education Goals Reports, Even Start parents are more likely to read to their children than parents nationally.

- ***Comparison to gains observed in other evaluations.*** Some studies compared local Even Start gains to gains from other studies of adult education or early childhood education programs. For example, one project compared Even Start parents with parents in a local JOBS program on expectations of their child and view of parenting. No differences were found between the two groups of parents. Two projects found that the CASAS gains of their adults were greater than the gains of adults in California adult education programs. Yet another program compared the CASAS gains of their adults to those of adults in a Toyota-funded family literacy project. Finally, one program found that the PSI scores of Even Start children were lower than the scores of children in a Home Start evaluation conducted 20 years ago.

- ***Comparison to "normal" development.*** Some studies compared local Even Start gains to normal development based on the use of local or

national norms. For example, one project found large gains when comparing developmental age with chronological age using PPVT scores; a second project projected normal growth based on pretest scores of the Learning Accomplishment Profile and found that Even Start children gained more than expected.

These local evaluations made an important attempt to judge the magnitude of the gains observed for Even Start children or adults. However, none of the studies was able to estimate the "true" impact of Even Start, which would require randomly assigning incoming families to Even Start or to a control group. Given the resources available to local evaluators, it is unreasonable to expect that any local study could actually put a randomized experiment in place.

Participant Characteristics Associated with Positive Outcomes

Only a scattering of local evaluations provided any information on the way in which participant characteristics are associated with outcomes. Local evaluators do not provide such breakdowns because few local evaluations have sufficiently large sample sizes to do any but the most aggregated analyses. Sample sizes in the studies we reviewed were most often in the range of 15 to 50 families. Therefore, while some of the raw score differences between the groups compared appear to be relatively large, they are not statistically significant because the sample sizes are so small, often fewer than 10 per group.

One study examined child gains separately for boys and girls, and for whites and Hispanics, with no differences between groups. A second study analyzed TABE reading, math, and language gains for adults who participated for over/under 150 hours and showed greater gains for the more active participants. A third study analyzed data separately for "new" vs. "continuing" children, younger vs. older preschool classrooms, Spanish vs. English test administration, and boys vs. girls, with no consistent differences for any comparison. Finally, a fourth study analyzed PPVT scores by age, with no differences in the gains of younger vs. older children.

6.0 Recommendations

The Department currently spends much more on local Even Start evaluations (more than $3 million per year) than on the national evaluation ($500,000 to $750,000 per year). If the mere existence of a local evaluation report is useful, or if it is useful for project directors to have a local evaluation that concludes that Even Start is a success, then local evaluations can be judged helpful to local projects. However, if the criterion for success is that a local evaluation should be used to help improve an Even Start project, then this review of more than 100 evaluations was not able to show that local studies provide much useful information, either to local project directors, to State Coordinators, or to the Department.

How can we improve this situation? Most important, in our view, is to provide clear guidance on the purpose of local evaluations. The current mix of guidance (to assess program outcomes, to address questions of local importance, to improve program functioning) needs to be streamlined and focused. We agree with the importance of the Department's recent move towards a modification of the current system in which local grantees would be encouraged to use local evaluations as vehicles for program improvement. Some specific recommendations for improving local Even Start evaluations and for enhancing the ability of Even Start grantees to conduct continuous improvement efforts are given below.

Change the Even Start legislation and preliminary guidance to refer to "local continuous improvement efforts" instead of "local evaluations." Terminology is important, and the past and current use of the term "local evaluation" has become synonymous with a study only of program outcomes, with the implication that local projects will be in trouble if they cannot demonstrate positive outcomes. Use of the term "local continuous improvement effort" would more clearly signal the understanding that Even Start is a difficult program to implement, and the intention that projects are expected to engage in a systematic, ongoing assessment of the strengths and weaknesses, as well as the outcomes, of their approaches, and to use evaluation data to improve their programs.

Provide guidance to state coordinators and/or local projects on the amount of funding that needs to be spent on a good local evaluation which focuses on program improvement. Without such guidance new projects have no idea of the amount of funds to allocate for this activity, and some do not include a local evaluation line item in their budgets. Suggesting a set amount or a percentage of the total project's budget that should be allocated to evaluation will signal the importance of evaluation and data collection to local program staff.

Help establish a community of local evaluators that can exchange information about useful evaluation approaches by making a mailing list of local evaluators available, by helping local evaluators exchange e-mail addresses, and by having sessions for local evaluators at annual Even Start conferences.

Provide training for local grantees in using data collected at the state and/or national level for continuous improvement at the local level. Local family literacy projects often regard participation in state-level and national-level studies as a burden that offers little or no return. State and national evaluation sponsors could help ensure local use of data by providing training to local staff in (1) accessing data collected for state or national studies, (2) understanding what those data mean at the local level, and (3) using those data to assess the performance level of their projects and to improve unsatisfactory performance.

Provide guidance to local projects on the use of data/evaluation for program improvement. Program improvement currently occurs through trial-and-error, and is based on the hard-learned experiences of program staff. This is one way of improving programs, but greater gains could be made by accompanying these personal, anecdotal methods with a data-based systematic assessment of program strengths and weaknesses. Such systematic continuous improvement efforts have been described by Haslam (1998) in the Observational Study of Even Start Projects, by Alamprese (1996) in studies of workplace literacy programs, and by Appel (1998) in her work with local Even Start evaluations.

The Department, in the Observational Study of 10 well-implemented Even Start projects, is facilitating and assessing the use of the continuous improvement approach based on the following principles:

Collaboratively set outcome goals for children and families. Local project staff, evaluators, and families enrolled in the program work together to set concrete outcome goals to be achieved.

Devise an intervention to achieve those goals. The family literacy intervention should be designed to achieve the specific outcome goals set in the first step. The intervention should be based on practices shown to be effective in prior research as well as on proven program quality indicators.

Set intervention thresholds necessary to achieve goals. An assumption underlying all family literacy programs is that if a family is to achieve their goals they must participate in the program for a sufficiently long period of time at a sufficiently high level of intensity. Program staff need to define, ahead of time, the minimum intervention threshold that they believe is needed for families to achieve their goals.

Assess progress toward goals with sound measures. Assessing progress on a periodic basis by using sound measures (i.e., measures that have adequate reliability and validity, that have a history of use in similar studies, that are available in appropriate languages, and so on) is the best way of determining whether families have achieved their goals.

Use evaluation to monitor program quality and results, and to target areas for improvement. A comprehensive evaluation will help program staff monitor the quality of their family literacy program. It also will help staff keep track of each family's level and duration of participation in the intervention and assess progress towards goals. Knowing the quality of each program component, the extent to which families have participated, and the degree to which they are making progress will allow program staff to understand why some families do not achieve their goals (e.g., they did not participate at a sufficient level, or program components were of low quality), to identify program components that need improvement, and to target resources accordingly.

7.0 References

Alamprese, J. (1996). The Role of Data in Program Management and Accountability in Adult Education. In Tana Reiff (ed.) The Pennsylvania ABLE Administrators Handbook, 43-44. Harrisburg: Pennsylvania Department of Education, Bureau of Adult Basic and Literacy Education.

Appel, E. (1998). National Even Start Evaluation Conference: Keynote Address. Baltimore, MD; Houston, TX; Denver, CO.

Cooper, H. & Hedges, L. (Eds.) (1994). The Handbook of Research Synthesis. New York: Russell Sage Foundation.

Dwyer, C. (1998). Local Evaluation User's Guide. Portsmouth, NH: RMC Research Corporation.

Haslam, B. (1998). Observational Study of Even Start Projects. Washington, DC: Policy Studies Associates.

McNaughton, D. (1994). Measuring parent satisfaction with early childhood intervention programs: Current practice, problems and future perspectives. Topics in Early Childhood Special Education, 14, 26-48

St.Pierre, R., Swartz, J., Gamse, B., Murray, S., Deck, D. & Nickel, P. (1995) National Evaluation of the Even Start Family Literacy Program: Final Report. Cambridge, MA: Abt Associates Inc.

Tao, F., Gamse, B. & Tarr, H. (February 1998). National Evaluation of the Even Start Family Literacy Program: 1994-97 Final Report. Alexandria, VA: Fu Associates, Ltd.